Extended Conflict-free Set 再考(モデル表現とその構築に関する理論と実際の研究)

勝野, 裕文

数理解析研究所講究録
495巻 1983年 131-143

Extended Conflict-free Set 再考

勝野 裕文 (Hirofumi Katsuno)
(電電公社 武蔵野通研)

1. はじめに

データベースの分野ではデータ従属性に関する理論が盛んに研究されている。ここでいうデータ従属性の定義は広い意味から狭い意味までさまざまに使われている。しかし、関数従属(FD),多値従属(MVD),結合従属(JD)がデータ従属性の基本概念であることに対する異論はない。ここで、FD, MVD, JDは次のような直観的な意味を持つ.

FD — 関数関係(表に依存しない概念)

MVD — 表rを射影によってr_1とr_2に分解した時に、$r = r_1 \bowtie r_2$となる為の必要十分条件($\bowtie$は結合を示す)

JD — 表が射影によってn個に分解($r = r_1 \bowtie r_2 \bowtie \cdots \bowtie r_n$)できる為の必要十分条件

このように一つのMVD又はJDは表の世界において明確な意味を持つ. しかし、表の世界を離れた現実世界における

MVD, JDの意味，特にMVD, JDからなる集合の意味が不明確な場合がある。意味の不明確なMVD集合を排除する為に，conflict-freeと呼ばれるMVD集合のクラスが考えられている。[8]

我々はconflict-free MVD集合にFDの概念をとりこむ為に，extended conflict-freeという概念を既に提案した[6]。本稿では，そのextended conflict-freeの直観的な意味とその問題点について述べる。尚，関係データベースの基本概念，用語は既知であると仮定する。本稿で未定義な概念，用語については[9]を参照すること。

2. MVD, JDの現実世界における意味

一つのMVD, JDに注目すると、その現実世界における意味はかなり明確になる。初めに，MVD, JDを定義する。

［定義］

属性集合Uの部分集合 $X_1, \ldots, X_n$ が $U = \bigcup_{i=1}^{n} X_i$ をみたす時に，関係 $r(U)$ において、結合従属（JD）$\bowtie[X_1, \ldots, X_n]$ が成り立つとは，

$$r(U) = r[X_1] \bowtie \cdots \bowtie r[X_n]$$

が成り立つことをいう。但し，$r[X_i]$ は $r(U)$ の X_i 上への射影を示す。

上の定義において、$n=2$の場合、$r(U)$においてMVDが成り立つといい、$\bowtie[X_1, X_2]$のかわりに、

$X_1 \cap X_2 \twoheadrightarrow X_1 - (X_1 \cap X_2)$（又は $X_1 \cap X_2 \twoheadrightarrow X_2 - (X_1 \cap X_2)$）

と書く。尚，左辺が共通なMVDはまとめて，$X \twoheadrightarrow Y_1, \cdots, X \twoheadrightarrow Y_n$ を $X \twoheadrightarrow Y_1 | \cdots | Y_n$ と書く。

一つのJD(MVD)は次の場合に成り立つ。

(☆) 属性集合 X_i に関する一つの関連が存在して，その関連のデータが関係 $r_i(X_i)$ で表わされている時、

$$r = r_1(X_1) \bowtie \cdots \bowtie r_n(X_n)$$

とすると，r において JD; $\bowtie[X_1, \cdots, X_n]$ が成り立つ。

[例1]

学生(S)がどの講義(C)を履習しているかを示す関連に関するデータを $r_1(S, C)$ で表わし、先生(T)がどの講義(C)を教えているかを $r_2(T, C)$ で，講義(C)でどの本(B)が教材となっているかを $r_3(B, C)$ で、学生(S)の専攻(M)と学年(Y)を $r_4(S, M, Y)$ でそれぞれ表わすとする。この時，$r = r_1 \bowtie r_2 \bowtie r_3 \bowtie r_4$ とすると，r において $\bowtie[SC, TC, BC, SMY]$ が成り立つ。

従って，r において JD;$\bowtie[X_1, \cdots, X_n]$ が成り立つことの一

つの意味は、異なる関連がn個あって、rはそれらの関連のデータを一つの関係で表わしているという点である。しかし(☆)で構成したrにおいては、

(GC) 任意のiに対して, $r[X_i]=r_i$ (GC: globally consistent)

が成り立つとは限らない。即ち、(☆)のようにrを構成すると情報が失なわれる可能性がある。一般に, pairwise consistent を

(PC) 任意のi, j $(i \neq j)$に対して $r_i[X_i \cap X_j]=r_j[X_i \cap X_j]$

とすると、$n=2$であれば、(PC)が成り立てば、常に(GC)が成立する。一方, $n \geq 3$の時は、(PC)が成り立ったとしても, (GC)が成り立つとは限らない。

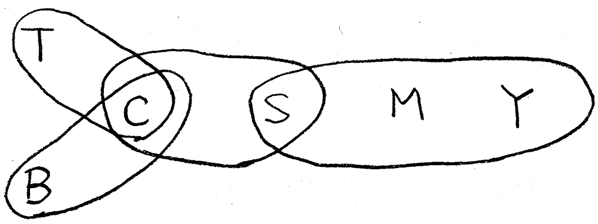
JDとMVDとの間の関係は単にMVDがJDの特殊な場合というだけではなく、一つのJDがMVDの集合と等価な場合がある。これが、MVDの集合の一つの意味である。

あるJDから導けるMVDの集合は次のように規定される。但し, "導ける"とは, $\bowtie \mathbb{R}$が成り立つすべての関係において、$X \twoheadrightarrow Y$が成り立つ時, $\bowtie \mathbb{R}$から$X \twoheadrightarrow Y$が導けるという。

$\mathbb{R}=\{X_1, \cdots, X_n\}$に対して, $\bowtie \mathbb{R}$によって$\bowtie[X_1, \cdots, X_n]$というJDを示すことにする。$\bowtie \mathbb{R}$に対応するハイパー・グラフとは各$X_i$の要素を頂点, X_iを辺とするハイパー・グラフをいう。

[例2]

例1のJDに対応するハイパー・グラフを下に示す。



定理１[5]

$\bowtie R$から$X \twoheadrightarrow Y$が導ける。

⇔ $\bowtie R$に対応するハイパー・グラフにおいて，Xを除去して得られるグラフの連結成分を$Y_1, \cdots, Y_m$とすると，ある$i_1, \cdots, i_\ell$が存在して，$Y = Y_{i_1} \cup \cdots \cup Y_{i_\ell}$とかける。

［例３］

例１のJD；$\bowtie$[SC, TC, BC, SMY]から例えば，

$CS \twoheadrightarrow T|B|MY$,　$S \twoheadrightarrow CTB|MY$

が導ける。

逆に，MVDの集合からJDが導ける場合，即ちMVDの集合とJDが等価になる場合は次の節で示す。

３．Conflict-free MVD集合

一つのMVD，JDの意味は２節でみたように比較的わかり易い。しかしMVDからなるある種の集合（MVDの推移律が適用できたり，MVDの左辺が他のMVDによって分割されるような集合）には意味的な問題点があることが示されている。[7][8] このような意味的な問題点が生じないMVDの集合の

クラスとして，conflict-freeという概念が定義されている。[8]

[定義]

ΓをFDとMVDからなる集合とする時，DEP(X)が次の条件をみたす時、DEP(X)をΓにおけるXの従属基であるという。

(1) $X=U$ ならば $DEP(X)=\{\phi\}$

(2) XのΓによる閉包をX^*とする時，$X^*=U$ならば、

$DEP(X)=\{A_1,\cdots,A_m \parallel \phi\}$ かつ

$U-X=\{A_1,\cdots,A_m\}$

(1),(2)以外の場合は次の(3)〜(6)が成り立つ。

(3) $DEP(X)=\{A_1,\cdots,A_m \parallel Y_1,\cdots,Y_k\}$ であり、$\{A_1,\cdots,A_m,Y_1,\cdots,Y_k\}$はU−Xの分割である。

(4) Γから $X\rightarrow A_i$ $(1\leqq i\leqq m)$ が導けるが、$B\in Y_1\cup\cdots\cup Y_k$ であれば、Γから $X\rightarrow B$ は導けない。i.e. $X^*=X\cup\{A_1,\cdots,A_m\}$

(5) Γから $X\twoheadrightarrow Y_i$ $(1\leqq i\leqq k)$ が導ける。

(6) Yの任意の空でない部分集合Y'に対して，Γから $X\twoheadrightarrow Y'$ は導けない。

尚，ΓがMVDのみからなる場合又は$X^*=X$の場合は，DEP(X)を単に$\{Y_1,\cdots,Y_k\}$と書く。

[定義]

MVDからなる集合をMとし、X,YをそれぞれMのあるM

VDの左辺であるとする時，XとYがconflict-freeであるとは，次の二条件が成り立つことをいう。

① $DEP(X) = \{V_1, \cdots, V_k, X_1, \cdots, X_i, Z_a \cup Y_1 \cup \cdots \cup Y_j\}$
$DEP(Y) = \{V_1, \cdots, V_k, Y_1, \cdots, Y_j, Z_b \cup X_1 \cup \cdots \cup X_i\}$
かつ $Z_a \cup X = Z_b \cup Y$ とかける．

② $\{V_1, \cdots, V_k\} \subset DEP(X \cap Y)$

M内のMVDの左辺の任意のpair，X，Yがconflict-freeである時，Mがconflict-freeであるという。

JD：⋈R に対して，ハイパーグラフを考え，そのグラフにおいて，サイクルの有無を議論することができる。ハイパーグラフがサイクルを持たない時に，⋈Rをacyclic JDと呼び，conflict-free MVD集合との間に次の定理が成り立つ。

定理2 [2]

Mがconflict-free MVD集合である．
$\Longleftrightarrow$ Mと同値なacyclic JDが存在する．

ここでいうハイパー・グラフのサイクルの概念の定義 [1] は面倒なので，省略するが，ここではacyclic JDを特徴づける次の定理を示す。

定理3 [2]

$\bowtie R$ が acyclic JD である.

$\Longleftrightarrow$ ハイパー・グラフ R から導けるMVDの集合と $\bowtie R$ が同値である.

$\Longleftrightarrow$ (PC) が成り立てば、(GC) が成り立つ。

定理3で示した他にも、acyclic JD (又は acyclic hypergraph) がいくつかの望ましい特徴を持つことが示されている。

4. Extended Conflict-free

前節で定義した conflict-free はMVDの集合に対する概念であって、FDの存在は特に議論されていない。[6]では、FDとMVDを一緒に考えて、extended conflict-free という概念を次のように定義した。

[定義]

Γ をFDとMVDからなる集合とする時、Xの従属基が

$$DEP(X) = \{A_1, \cdots, A_n \,\|\, Y_1, \cdots, Y_k\}$$

であれば、Xの退化MVD集合 M_X を

$$M_X = \{X^* \twoheadrightarrow Y_1, \cdots, X^* \twoheadrightarrow Y_k\} \quad \text{但し} \quad X^* = X \cup \{A_1, \cdots, A_n\}$$

と定義する。Γ の退化MVD集合 M を

$$M = \bigcup_X M_X \quad (\text{但し} \bigcup_X \text{において、}X\text{は}\Gamma\text{の}FD\text{又は}MVD\text{の左辺})$$

と定義する.

[定義]

Γの退化MVD集合Mがconflict-freeである時、Γがextended conflict-freeであるという。

conflict-free MVD集合がacyclic JDと同値であるように、extended conflict-free集合もJDと次のような対応がある。

[定義]

$X \rightarrow Y$が$\bowtie[X_1, \cdots, X_n]$と適合しているとは、ある$i$に対して、$X \cup Y \subset X_i$が成り立つことをいう。FDからなる集合Fの任意の元が$\bowtie[X_1, \cdots, X_n]$と適合する時に、Fは$\bowtie[X_1, \cdots, X_n]$に適合しているという。

[定義]

Fが$\bowtie[X_1, \cdots, X_n]$に完全に適合しているとは、Fが$\bowtie[X_1, \cdots, X_n]$に適合していて、かつ$X \rightarrow Y$がFの元であり、$X \subset X_i$ならば、$Y \subset X_i$が成り立つことをいう。

[例4]

例1において、$F = \{C \rightarrow T, S \rightarrow MY\}$とすると、Fは$\bowtie[SC, TC, BC, SMY]$に適合しているが、完全には適合していない。

定理4

FDとMVDからなる集合Γがextended conflict-freeである.

$\Longleftrightarrow$ あるFDの集合 F と acyclic JD: $\bowtie R$ が存在して、F は $\bowtie R$ と適合していて、かつ $F \cup \{\bowtie R\}$ は Γ と同値である。

$\Longleftrightarrow$ あるFDの集合 F と acyclic JD: $\bowtie R$ が存在して、F は $\bowtie R$ と完全に適合していて、かつ $F \cup \{\bowtie R\}$ は Γ と同値である。

extended conflict-free という概念を定義する目的の一つは、現実世界の自然な制約をデータ従属性の集合で記述する時に、その集合がどのように特徴づけられるかを明らかにすることであった。その目的からみると、extended conflict-free には次の例に示すような問題点がある。即ち、現実世界の自然な制約で、extended conflict-free 集合に対応しない場合がある。

[例5][5]

学生(S)がどの講義(C)を履習しているかを示すデータに対応する関係 $r_1(S, C)$ と、講義(C)が何時(H)にどの教室(R)で行なわれるかを示す関係 $r_2(C, H, R)$ がある時に、データ従属性の集合 Γ として、

$$\Gamma = \{SH \to CR,\ RH \to C,\ CH \to R,\ \bowtie[CS, CRH]\}$$

を考える。この Γ に対応する現実世界の制約は自然なものであるが、Γ と同値な extended conflict-free 集合は存在しない。

extended conflict-freeはこのような問題点を含むが、望ましい性質も持っている。それは、制約をextended conflict-freeに限ると、次に示すような忠実な分解が存在するという点である。

［定義］

Γをデータ従属性の集合とする時，Γを制約として持つ、属性集合R上の関係スキーマを$\langle R, \Gamma \rangle$と書くことにする。

［定義］

次の(1)〜(4)の性質を持つ、関係スキーマの集合$\{\langle R_1, F_1 \rangle, \cdots, \langle R_n, F_n \rangle\}$を$\langle R, \Gamma \rangle$の忠実な分解という.

(1) 各F_i $(1 \leq i \leq n)$はFDの集合である.

(2) Γと$F_1 \cup \cdots \cup F_n \cup \{\bowtie[R_1, \cdots, R_n]\}$は同値である.

(3) Γから導けるすべてのFDの集合をFとし、FのR_i上への制限を$F|_{R_i}$とすると，$F|_{R_i} = (F_i)^+$が成り立つ.
但し，$(\cdot)^+$はFD集合の閉包を示す。

(4) 各$\langle R_i, F_i \rangle$は第3正規形である.

［注意］

上の条件(2)は次の(2)'と同値であることが知られている.[3]

(2)' $SAT(F_1, \cdots, F_n) = \{\langle r_1, \cdots, r_n \rangle \mid r_i$は$F_i$が成り立つ関係であり，$r_1, \cdots, r_n$で(PC)が成り立つ$\}$とし，

$SAT(\Gamma) = \{r \mid r$はΓが成り立つ関係$\}$,

$r \in SAT(\Gamma)$ に対して $\Delta(r) = \langle r[R_1], \cdots, r[R_n] \rangle$ とすると, Δ は $SAT(\Gamma)$ から $SAT(F_1, \cdots, F_n)$ への全単射であり、Δ^{-1} は結合である.

定理5

Γ が extended conflict-free ならば、$\langle R, \Gamma \rangle$ の忠実な分解が存在する。

extended conflict-free が忠実な分解に対応しているか、即ち $\langle R, \Gamma \rangle$ の忠実な分解が存在すれば、Γ が extended conflict-free 集合に同値になるかというと, それは成立しない。忠実な分解は、FD集合とJDの適合性に関係している.

定理6

Γ を $\{\bowtie R\}^{\cup} F$ とする時, F が $\bowtie R$ と適合しているならば, $\langle R, \Gamma \rangle$ の忠実な分解が存在する. 逆に, $\langle R, \Gamma \rangle$ の忠実な分解が存在すれば、あるJD ; $\bowtie R$ とFD集合 F で、$\{\bowtie R\}^{\cup} F$ が Γ と同値で、F が $\bowtie R$ と適合しているものが存在する。

[注意]

定理4と定理6から定理5は容易に導かれる.

以上のことから考えると直観的に

extended conflict-free ≅ acyclic + 忠実な分解
という見方をすることができる。

5. おわりに

extended conflict-freeでは、[1]でいうacyclicという概念を前提としているが、ハイパー・グラフのサイクルの概念は多様なので、[1]のサイクルと異なる意味のサイクル[4]を対象とした時の考察も必要だと考える。

参考文献

[1] C. Beeri, R. Fagin, D. Maier, A. Mendelzon, J. Ullman and M. Yannakakis: Properties of Acyclic Database Schemes, Proc. of ACM-STOC, pp.355-362, 1981
[2] C. Beeri, R. Fagin, D. Maier and M. Yannakakis: On the Desirability of Acyclic Database Schemes, IBM Research Report, RJ3131, 1981
[3] C. Beeri and J. Rissanen: Faithful Representations of Relational Database Schemes, IBM Research Report, RJ2722, 1980
[4] Fagin, R.: Types of Acyclicity for Hypergraphs and Reltional Database Schemes, IBM Research Report, RJ3330, 1981
[5] Fagin, R., A.O. Mendelzon, and J.D. Ullman: A Simplified Universal Relation Assumption and its Properties, ACMTODS, Vol.7, No.3, pp.343-360, 1982
[6] Katsuno, H.: A Desirable Class of Sets of Functional Dependencies and Multivalued Dependencies, AL81-65, 1981
[7] Kambayashi, Y., K. Tanaka, and S. Yajima: Semantic Aspects of Data Dependencies and their Application to Relational Database Design, Proc. of COMPSAC 79, pp.398-403, 1979
[8] E. Sciore: Real-World MVDs, Proc. of ACM-SIGMOD, pp.121-132, 1981
[9] J.D. Ullman: Principles of Database Systems, Computer Science Press, 1980